Convergent semi-Lagrangian methods for the Monge-Ampère equation on unstructured grids

Article (Published Version)

SIAM J. NUMER. ANAL.
Vol. 55, No. 2, pp. 691–712

CONVERGENT SEMI-LAGRANGIAN METHODS FOR THE MONGE–AMPÈRE EQUATION ON UNSTRUCTURED GRIDS*

XIAOBING FENG† AND MAX JENSEN‡

Abstract. This paper is concerned with developing and analyzing convergent semi-Lagrangian methods for the fully nonlinear elliptic Monge–Ampère equation on general triangular grids. This is done by establishing an equivalent (in the viscosity sense) Hamilton–Jacobi–Bellman formulation of the Monge–Ampère equation. A significant benefit of the reformulation is the removal of the convexity constraint from the admissible space as convexity becomes a built-in property of the new formulation. Moreover, this new approach allows one to tap the wealthy numerical methods, such as semi-Lagrangian schemes, for Hamilton–Jacobi–Bellman equations to solve Monge–Ampère-type equations. It is proved that the considered numerical methods are monotone, pointwise consistent, and uniformly stable. Consequently, its solutions converge uniformly to the unique convex viscosity solution of the Monge–Ampère Dirichlet problem. A superlinearly convergent Howard's algorithm, which is a Newton-type method, is utilized as the nonlinear solver to take advantage of the monotonicity of the scheme. Numerical experiments are also presented to gauge the performance of the proposed numerical method and the nonlinear solver.

Key words. Monge–Ampère equation, Hamilton–Jacobi–Bellman equation, viscosity solution, semi-Lagrangian method, wide stencil, monotone scheme, convergence, Howard's algorithm

AMS subject classifications. 65N06, 65N12, 65N35, 35J60

DOI. 10.1137/16M1061709

1. Introduction.

This paper is concerned with semi-Lagrangian methods for the following Dirichlet boundary value problem of a fully nonlinear elliptic Monge–Ampère-type equation:

$$\det(D^2u) = \Big(\frac{f}{d}\Big)^d \qquad \text{in } \Omega, \tag{1a}$$

$$u(x) = g(x) \qquad \text{on } \partial\Omega, \tag{1b}$$

where Ω and $\partial\Omega$ denote respectively a bounded strictly convex domain in $\mathbf{R}^d$ $(d \geq 2)$ and its boundary. The Hessian of the function u is denoted D^2u. The functions $f : \Omega \to [0, \infty)$ and $g : \partial\Omega \to \mathbf{R}$ are bounded and continuous. We note that the special form of the right-hand side in (1a) is chosen for notational convenience in the subsequent analysis; the usual form can be easily recovered by setting $f = d\widetilde{f}^{\frac{1}{d}}$.

Monge–Ampère-type equations, along with Hamilton–Jacobi–Bellman-type equations (see below), are two major classes of fully nonlinear second-order partial differential equations (PDEs). They arise from many scientific and technological applications such as antenna design, astrophysics, differential geometry, image processing, optimal mass transport, and semigeostrophic fluids, to name a few (see [15, section 5] for details). From the PDE point of view, Monge–Ampère-type equations are well understood; see [18, Chapter 17] for a detailed account on the classical solution theory

*Received by the editors February 16, 2016; accepted for publication (in revised form) December 5, 2016; published electronically March 16, 2017.
http://www.siam.org/journals/sinum/55-2/M106170.html
Funding: The work of the first author was partially supported by NSF grant DMS-0710831.
†Department of Mathematics, University of Tennessee, Knoxville, TN 37996 (xfeng@math.utk.edu).
‡Department of Mathematics, University of Sussex, Brighton BN1 9QH, United Kingdom (m.jensen@sussex.ac.uk).

and [19, 9] for the viscosity solution theory. On the other hand, from the numerical point of view, the situation is far from ideal. Very few numerical methods, which can reliably and efficiently approximate viscosity solutions of Monge–Ampère-type PDEs on general convex domains, are available in the literature (see [8, 15, 16, 17, 26, 29] and the references therein). There are three main difficulties which lead to the lack of progress on approximating viscosity solutions of fully nonlinear second-order PDEs. First, the fully nonlinear structure and nonvariational concept of viscosity solutions of the PDEs prevent a direct formulation of any Galerkin-type numerical methods (such as finite element, discontinuous Galerkin, and spectral methods). Second, the Monge–Ampère operator, $u \mapsto \det(D^2u)$, is not an elliptic operator in generality; instead, it is elliptic only in the set of convex functions and the uniqueness of viscosity solutions only holds in that space. This convexity constraint, imposed on the admissible space, causes a daunting challenge for constructing convergent numerical methods; it indeed screens out any trivial finite difference and finite element analysis because the set of convex finite element functions is not dense in the set of convex functions [2]. Third, as the right-hand side f of (1a) vanishes, the Monge–Ampère mapping attains characteristics of a degenerate elliptic operator. In this setting the regularity of exact solutions is reduced, limiting the tools available for a convergence analysis of numerical solutions.

The goal of this paper is to develop a new approach for constructing convergent numerical methods for the Monge–Ampère Dirichlet problem (1), in particular, by focusing on overcoming the second difficulty caused by the convexity constraint. The crux of the approach is to first establish an equivalent (in the viscosity sense) Bellman formulation of the Monge–Ampère equation and then to design monotone semi-Lagrangian methods for the resulting Bellman equation on general triangular grids. The proposed methods are closely related to two-grid constructions because we use a finite element ambient grid to define the approximation space, combined with wide finite difference stencils layered over this ambient grid. An aim in the design of the numerical schemes is to make Howard's algorithm available, which is a globally superlinearly converging semismooth Newton solver. This allows us to robustly compute numerical approximations on very fine meshes of nonsmooth viscosity solutions, including the degenerate case where $f \geq 0$. An advantage of the rigorous convergence analysis of the numerical solutions is the comparison principle for the Bellman operator, which extends to nonconvex functions. We deviate from the established Barles–Souganidis framework in the treatment of the boundary conditions to address challenges arising from consistency and comparison. The proposed approach also bridges the gap between advances on numerical methods for these two classes of second-order fully nonlinear PDEs; see, for instance, [6, 10, 11, 13, 21, 25, 30] and the references therein for the numerical literature on Bellman equations.

The remainder of this paper is organized as follows. In section 2 we collect preliminaries including the definition of viscosity solutions. In section 3 we introduce a well-known Hamilton–Jacobi–Bellman reformulation of the Monge–Ampère equation in the classical solution setting and prove such an equivalence still holds in the viscosity solution framework. In section 4 we introduce a numerical scheme (21) for the Monge–Ampère equation. In section 5 we prove the existence and uniqueness of numerical solutions and present a globally converging semismooth Newton method. Section 6 contains the main result of the paper: Theorem 6.5 demonstrates the uniform convergence to the unique viscosity solution. In section 7 we relate the class of schemes of this paper to existing methods to solve Hamilton–Jacobi–Bellman

equations. In section 8 we present numerical experiments which verify the accuracy and efficiency of the proposed method and the nonlinear solver.

2. Viscosity solutions. Let $\Omega \subset \mathbf{R}^d$ be a bounded open strictly convex domain. We denote by $B(G)$, $\mathrm{USC}(G)$, and $\mathrm{LSC}(G)$, respectively, the spaces of bounded, upper semicontinuous, and lower semicontinuous functions on a set $G \subset \mathbf{R}^d$. For any $v \in B(\overline{\Omega})$, we define

$$v^*(x) := \limsup_{y \to x} v(y) \qquad \text{and} \qquad v_*(x) := \liminf_{y \to x} v(y).$$

Then, $v^* \in \mathrm{USC}(\overline{\Omega})$ and $v_* \in \mathrm{LSC}(\overline{\Omega})$, and they are called the *upper* and *lower semicontinuous envelopes* of v, respectively.

Given a bounded function $F : \mathbf{S} \times \mathbf{R}^d \times \mathbf{R} \times \Omega \to \mathbf{R}$, where $\mathbf{S}$ denotes the set of $d \times d$ symmetric real matrices, the general second-order fully nonlinear PDE takes the form

$$F(D^2u, \nabla u, u, x) = 0 \qquad \text{in } \Omega. \tag{2}$$

We impose Dirichlet boundary conditions in the pointwise sense that $u(x) = g(x)$ for all $x \in \partial\Omega$. In the discussion about converging numerical schemes we shall draw comparisons with Dirichlet conditions in the viscosity sense, which are imposed as a discontinuity of the PDE; cf. [4, p. 274] and [12, section 7.C].

The following definitions can be found in [4, 9, 12, 18, 19].

DEFINITION 2.1. *A function $u \in \mathrm{USC}(\Omega)$ (resp., $u \in \mathrm{LSC}(\Omega)$) is called a viscosity subsolution (resp., supersolution) of* (2) *if for all $\varphi \in C^2(\Omega)$ such that $u - \varphi$ has a local maximum (resp., minimum) at $x \in \Omega$ we have*

$$F(D^2\varphi(x), \nabla\varphi(x), u(x), x) \le 0$$

(resp., $F(D^2\varphi(x), \nabla\varphi(x), u(x), x) \ge 0$). The function u is said to be a viscosity solution of (2) *if it is simultaneously a viscosity subsolution and supersolution of* (2).

The restriction to convex functions in Definition 2.2 below reflects that the Monge–Ampère equation is elliptic only on the set of convex functions, while the Hamilton–Jacobi–Bellman operator of our subsequent construction is elliptic on the whole space. For details we refer to [19, section 1.3].

DEFINITION 2.2. *A function $u \in \mathrm{USC}(\Omega)$ (resp., $u \in \mathrm{LSC}(\Omega)$) is called a viscosity subsolution (resp., supersolution) of* (2) *on the set of convex functions if u is convex and if for all convex $\varphi \in C^2(\Omega)$ such that $u - \varphi$ has a local maximum (resp., minimum) at $x \in \Omega$ we have*

$$F(D^2\varphi(x), \nabla\varphi(x), u(x), x) \le 0$$

(resp., $F(D^2\varphi(x), \nabla\varphi(x), u(x), x) \ge 0$). The function u is said to be a viscosity solution of (2) *on the set of convex functions if it is simultaneously a viscosity subsolution and supersolution of* (2) *on the set of convex functions.*

Note that in Definition 2.2 the set of test functions is smaller. Therefore it is not obvious that viscosity solutions on the set of convex functions are solutions in the sense of Definition 2.1.

3. Hamilton–Jacobi–Bellman form of the Monge–Ampère equation. It is known [23, 27] that the Monge–Ampère equation has an equivalent Hamilton–Jacobi–Bellman (or Bellman for brevity) formulation in the setting of classical solutions. However, to the best of our knowledge, such an equivalence has not been

extended to the case of viscosity solutions in the literature. The goal of this section is to prove this extension rigorously. A related description of the relationship between classical and viscosity solutions is examined in terms of elliptic sets in [24].

Let $\mathbf{S}_+ := \{A \in \mathbf{S};\, A \geq 0\}$ and $\mathbf{S}_1 := \{B \in \mathbf{S}_+;\, \operatorname{tr} B = 1\}$. It is easy to check [23] that $\mathbf{S}_1$ is a compact subset of $\mathbf{S}_+$ and, consequently, $\mathbf{S}_1$ is bounded in the Euclidean norm.

We define the Bellman operator

$$H(A,f) := \sup_{B\in\mathbf{S}_1} \left(-B : A + f\sqrt[d]{\det B}\right) \qquad \forall A \in \mathbf{S}, f \in [0,\infty) \tag{3}$$

and the Monge–Ampère operator

$$M(A,f) := \left(\frac{f}{d}\right)^d - \det(A) \qquad \forall A \in \mathbf{S}, f \in [0,\infty). \tag{4}$$

Then the Monge–Ampère problem (1) can be rewritten as

$$M\big(D^2u(x), f(x)\big) = 0 \qquad \forall x \in \Omega, \tag{5a}$$

$$u(x) = g(x) \qquad \forall x \in \partial\Omega, \tag{5b}$$

which gives the structure of (2) upon setting

$$F\big(D^2u(x), \nabla u(x), u(x), x\big) = M\big(D^2u(x), f(x)\big).$$

Analogously we also define the Bellman problem

$$H\big(D^2u(x), f(x)\big) = 0 \qquad \forall x \in \Omega, \tag{6a}$$

$$u(x) = g(x) \qquad \forall x \in \partial\Omega \tag{6b}$$

with the correspondence $F\big(D^2u(x), \nabla u(x), u(x), x\big) = H\big(D^2u(x), f(x)\big)$.

The proofs of the following Lemmas 3.1 and 3.2 are given in [23, p. 51].

LEMMA 3.1. *There exists a maximizer $B' \in \mathbf{S}_1$ of the supremum in* (3) *which commutes with $A \in \mathbf{S}$. In particular, there is a coordinate transformation, depending on A, which simultaneously diagonalizes A and B'.*

The next result gives equivalence of convex classical solutions of (5) and (6). We highlight that the lemma covers the degenerate case $f = 0$.

LEMMA 3.2. *Let $f \in [0,\infty)$ and $A \in \mathbf{S}$. Then $H(A,f) = 0$ holds if and only if $M(A,f) = 0$ and $A \in \mathbf{S}_+$.*

We remark that there is another slightly different Bellman reformulation of the Monge–Ampère problem (5) which uses a determinant constraint (instead of a trace constraint) on the control B in the definition of the Hamiltonian H; see [27]. However, the numerical discretization of a determinant constraint is less straightforward, explaining our preference for (3).

Let D_ℓ be the matrix $(\delta_{i\ell}\delta_{j\ell})_{ij}$ which vanishes in all entries except for the ℓth diagonal term which is 1.

THEOREM 3.3. *Let $f \in C(\Omega)$ be nonnegative and u be a viscosity subsolution (supersolution) of the Monge–Ampère problem* (5a) *on the set of convex functions. Then u is a viscosity subsolution (supersolution) of Bellman problem* (6a).

Proof. Step 1. We first consider the case that u is a viscosity subsolution of (5a). Let $\phi \in C^2(\Omega)$ such that $u - \phi$ attains a local maximum at $x \in \Omega$. Since u is convex it follows that ϕ is convex in a neighborhood N of x; cf. [19, Remark 1.3.2]. By the definition of viscosity subsolutions on the set of convex functions, noting the local character of the definition, we have $M\big(D^2\phi(x), f(x)\big) \le 0$.

Let $\xi \ge 0$ such that $M(D^2\phi(x), f(x)) + \xi = 0$. Equivalently,

$$M(D^2\phi(x), \widehat{f}) = 0 \quad \text{with} \quad \widehat{f} := d\sqrt[d]{\Big(\frac{f(x)}{d}\Big)^d + \xi} \ge f(x).$$

By Lemma 3.2 we have $H(D^2\phi(x), \widehat{f}) = 0$. Thus, u is a viscosity subsolution of (6a), using that $g \mapsto H(D^2\phi(x), g)$ is monotonically increasing.

Step 2. Now we consider the case that u is a viscosity supersolution of (5a). The proof of this step differs because now nonconvex ϕ which are test functions for H but not M need to be considered and because a negative slack variable ξ can in general not be covered by Lemma 3.2.

Let $\phi \in C^2(\Omega)$ such that $u - \phi$ attains a local minimum at $x \in \Omega$.

(a) We first suppose that ϕ is convex in a neighborhood of x. Then we have $M\big(D^2\phi(x), f(x)\big) \ge 0$ and that

$$\Big(\frac{f}{d}\Big)^d \ge \det(D^2\phi(x)) \ge 0.$$

Hence with $\widehat{f} := d\sqrt[d]{\det(D^2\phi(x))}$ there holds $f(x) \ge \widehat{f} \ge 0$ and $M(D^2\phi(x), \widehat{f}) = 0$. Due to Lemma 3.2, $H(D^2\phi(x), f(x)) \ge H(D^2\phi(x), \widehat{f}) = 0$.

(b) Now suppose that ϕ is not convex in the vicinity of x. We may assume without loss of generality that $D^2\phi(x)$ is diagonal. Then there is a $\partial^2_{\ell\ell}\phi(x) \le 0$. Therefore

$$H(D^2\phi(x), f(x)) \ge -D_\ell : D^2\phi(x) = -\partial^2_{\ell\ell}\phi(x) \ge 0.$$

Parts (a) and (b) guarantee that u is a viscosity supersolution of (6a). ∎

To show that solutions of the Bellman problem solve the Monge–Ampère problem, convexity needs to be enforced. We first prove a technical lemma.

LEMMA 3.4. *Let $A \in \mathbf{S}_+$, $f \in [0, \infty)$ and let λ be the smallest eigenvalue of A. Then the function*

$$\Phi_{A,f} : \ [-f, \infty) \to [-\lambda, \infty), \ \delta \mapsto H(A, f + \delta)$$

is continuous, strictly monotonically increasing, and bijective.

Proof. We assume without loss of generality that A is a diagonal matrix and that λ is the first entry on the diagonal of A.

If $\delta = -f$, then the function value of $H(A, f + \delta)$ cannot be affected by the term $(f + \delta)\sqrt[d]{\det B}$ in (3) for any $B \in \mathbf{S}_1$. Hence $D_1 \in \mathbf{S}_1$ is a maximizer in (3) and $H(A, f + \delta) = -\lambda$.

Now let $\delta > -f$ and consider $B_\alpha = \alpha \operatorname{Id} + (1 - d\alpha)D_1$. Then, as $\alpha \to 0$,

$$-B_\alpha : A = -\alpha \operatorname{tr} A - (1 - d\alpha)\lambda = -\lambda + \mathcal{O}(\alpha).$$

Similarly,

$$\sqrt[d]{\det B_\alpha}(f + \delta) = \big((1 - (d-1)\alpha)\alpha^{d-1}\big)^{\frac{1}{d}}(f + \delta) = \mathcal{O}(\alpha^{1-1/d}).$$

It follows that there is an $\alpha \in (0,1]$ such that

$$-B_\alpha : A + (f+\delta)\sqrt[d]{\det B_\alpha} > -D_1 : A + (f+\delta)\sqrt[d]{\det D_1} = -\lambda.$$

As D_1 is maximizer over the set of singular matrices in $\mathbf{S}_1$, it is clear that the maximizer B' over all of $\mathbf{S}_1$ is invertible. Let $h > 0$. Then,

$$\Phi_{A,f}(\delta) < -B' : A + \sqrt[d]{\det B'}(f+\delta+h) \le H(A, f+\delta+h) = \Phi_{A,f}(\delta+h).$$

Hence $\Phi_{A,f}$ is strictly monotone and thus injective.

As supremum of affine functions, $\Phi_{A,f}$ is convex and therefore continuous. This with $\Phi_{A,f}(\delta) \ge \frac{1}{d}(f+\delta-\mathrm{tr}A)$, owing to the control $\frac{1}{d}\mathrm{Id} \in \mathbf{S}_1$, ensures that $\Phi_{A,f}$ is surjective. ∎

With Lemma 3.4 we can find for each A a suitable $\widehat{f}$ such that $H(A,\widehat{f}) = 0$.

THEOREM 3.5. *Let $f \in C(\Omega)$ be nonnegative and u be a viscosity solution of the Bellman problem* (6a)*. Then u is a viscosity solution of Monge–Ampère problem* (5a) *on the set of convex functions.*

Proof. Step 0. Let $x \in \Omega$ and let (p, A) belong to the second-order superjet

$$J^{2,+}u(x) := \left\{(D\phi(x), D^2\phi(x)) : \phi \in C^2 \text{ and } u-\phi \text{ has local maximum at } x\right\}. \tag{7}$$

Then

$$\sup_{B\in\mathbf{S}_1}\left(-B : A + f\sqrt[d]{\det B}\right) \le 0$$

due to the definition of viscosity subsolutions in terms of second-order jets instead of test functions. Thus $B : A \ge f\sqrt[d]{\det B} \ge 0$ for all $B \in \mathbf{S}_1$, implying that $A \ge 0$. It follows from [3, Lemma 1] that u is convex on Ω.

Step 1. We now show that u is a viscosity subsolution of (5a). Let $\phi \in C^2(\Omega)$ be convex such that $u-\phi$ attains a local maximum at $x \in \Omega$. Then $H\big(D^2\phi(x), f(x)\big) \le 0$. Let

$$\widehat{f} = f(x) + \Phi^{-1}_{D^2\phi(x),f(x)}(0)$$

so that $H(D^2\phi(x), \widehat{f}) = 0$. Since $H(D^2\phi(x), f(x)) \le 0$ it follows from monotonicity that $\widehat{f} \ge f(x) \ge 0$. By Lemma 3.2 we have $M(D^2\phi(x), \widehat{f}) = 0$. Thus, u is a viscosity subsolution of (5a).

Step 2. Now we show that u is a viscosity supersolution of (5a). Let $\phi \in C^2(\Omega)$ be convex such that $u-\phi$ attains a local minimum at $x \in \Omega$. Then we have $H\big(D^2\phi(x), f(x)\big) \ge 0$. Since $D^2\phi(x)$ is positive semidefinite we know that $H(D^2\phi(x), 0) \le 0$. So 0 is in the domain of $\Phi^{-1}_{D^2\phi(x),f(x)}$. Set

$$\widehat{f} = f(x) + \Phi^{-1}_{D^2\phi(x),f(x)}(0).$$

It follows $f(x) \ge \widehat{f} \ge 0$. By Lemma 3.2 we have $M(D^2\phi(x), \widehat{f}) = 0$. Thus, u is a viscosity supersolution of (5a). ∎

At this point we have shown that the set of viscosity solutions of the Bellman and Monge–Ampère operators coincides without imposing any boundary conditions. It is clear that the solution sets also coincide if Dirichlet conditions are enforced pointwise:

$$\begin{aligned}&\{v \in C(\overline{\Omega}) : \text{ viscosity solution of (5a)}\} \cap \{v \in C(\overline{\Omega}) : v|_{\partial\Omega} = 0\}\\ =&\{v \in C(\overline{\Omega}) : \text{ viscosity solution of (6a)}\} \cap \{v \in C(\overline{\Omega}) : v|_{\partial\Omega} = 0\}.\end{aligned}$$

We now turn to a comparison principle for the Bellman problem, which holds on the whole function space. This is an advantage over comparison principles for the Monge–Ampère problem, which are usually formulated for the set of convex functions.

LEMMA 3.6. *Let $u \in \mathrm{USC}(\overline{\Omega})$ be a subsolution and $v \in \mathrm{LSC}(\overline{\Omega})$ be a supersolution of the Bellman problem* (6a). *Then $u \le v$ on $\overline{\Omega}$ if $u \le v$ on $\partial\Omega$.*

Proof. We briefly outline how the comparison argument of section 5.C in [12] applies in this context. Suppose that $u \le v$ on $\partial\Omega$ but $u(x') > v(x')$ for some $x' \in \Omega$. For $\epsilon > 0$ set $u_\epsilon(x) := u(x) + \frac{\epsilon}{2}|x - x'|^2 - \frac{\epsilon}{2}\sup_{y\in\Omega}|y - x'|^2$, where $|\cdot|$ denotes the Euclidean norm. Notice that $u_\epsilon \le v$ on $\partial\Omega$. Moreover, for $x \in \Omega$, one has [12, Remark 2.7(ii)]

$$(p, X) \in \overline{J}^{2,+}u_\epsilon(x) \qquad \text{if and only if} \qquad (p - \nabla_x \tfrac{\epsilon}{2}|x - x'|^2, X - \epsilon\mathrm{Id}) \in \overline{J}^{2,+}u(x),$$

where we referred to the closures

$$\begin{aligned}\overline{J}^{2,+}u(x) := \big\{&(p, X) \in \mathbf{R}^d \times \mathbf{S} : \exists\, (x_n, p_n, X_n) \in \Omega \times \mathbf{R} \times \mathbf{S} \text{ so that}\\ &(p_n, X_n) \in J^{2,+}u(x_n) \text{ and } (x_n, u(x_n), p_n, X_n) \to (x, u(x), p, X)\big\}\end{aligned}$$

of the superjets (7) as required by Theorem 3.2 of [12] used below.

Now, with the maximizer B',

$$\begin{aligned}H(X, f(x)) &= \sup_{B\in\mathbf{S}_1}\Big(-B : X + f(x)\sqrt[d]{\det B}\Big) = -B' : X + f(x)\sqrt[d]{\det B'}\\ &= -B' : (X - \epsilon\mathrm{Id}) + f(x)\sqrt[d]{\det B'} - \epsilon \le H(X - \epsilon\mathrm{Id}, f(x)) - \epsilon \le -\epsilon,\end{aligned}$$

where we used that $B' : \mathrm{Id} = \mathrm{tr}\, B' = 1$.

We assume $\epsilon \in (0, 2(u(x') - v(x'))/\operatorname{diam}(\Omega)^2)$ because then $u_\epsilon(x') > v(x')$. Arguing with Proposition 3.7 of [12], for α sufficiently large there exist $(x_\alpha, y_\alpha) \in \Omega \times \Omega$ which maximize $(x, y) \mapsto u_\epsilon(x) - v(y) - \frac{\alpha}{2}|x - y|^2$, as the maxima cannot be attained at the boundary. Appealing to Theorem 3.2, (3.9), and (3.10) of [12], there are

$$(\alpha(x_\alpha - y_\alpha), X) \in \overline{J}^{2,+}u_\epsilon(x_\alpha), \qquad (\alpha(x_\alpha - y_\alpha), Y) \in \overline{J}^{2,-}v(y_\alpha)$$

such that $X \le Y$. Therefore

$$\begin{aligned}0 &= H(X, f(x_\alpha)) - H(Y, f(y_\alpha)) + H(Y, f(y_\alpha)) - H(X, f(x_\alpha))\\ &\le \tfrac{f(y_\alpha) - f(x_\alpha)}{d} - \epsilon,\end{aligned} \tag{8}$$

where we used $H(X, f(x_\alpha)) \le -\varepsilon$ and $H(Y, f(y_\alpha)) \ge 0$ and

$$\begin{aligned}H(Y, f(y_\alpha)) - H(X, f(x_\alpha)) &\le \sup_{B\in\mathbf{S}_1}\Big(-B : (Y - X) + (f(y_\alpha) - f(x_\alpha))\sqrt[d]{\det B}\Big)\\ &\le (f(y_\alpha) - f(x_\alpha))\sup_{B\in\mathbf{S}_1}\sqrt[d]{\det B}.\end{aligned}$$

Owing to the continuity of f we find $f(y_\alpha) - f(x_\alpha) \to 0$ as $\alpha \to \infty$, so that (8) is a contradiction. Hence $u_\epsilon(x) \le v(x)$ for small $\epsilon > 0$ and $x \in \overline{\Omega}$. ∎

Remark 3.7 (general boundary conditions and convexity). It is a straightforward exercise to show that we can impose the more general (possibly nonlinear) boundary conditions $(p, r, x) \to B(p, r, x)$ in the viscosity sense in (5) and (6), where the new argument p takes the role of a gradient, and retain equal solution sets.

We also observe that the proof of equivalence does not need convexity of the domain Ω. We note, however, a close relationship between boundary conditions, comparison, and convexity in [19] and also in section 6 below, where we study convergence of numerical methods.

4. Monotone semi-Lagrangian methods. In section 3 we prove that the Monge–Ampère problem (5) has a Bellman reformulation (6) in the viscosity sense. This equivalence opens a route for developing numerical methods for (5) via (6). There are major advantages in pursuing this approach.

(a) In (5) convexity is built into the boundary value problem as a constraint (cf. Definition 2.2) that is difficult to maintain at the discrete level. In contrast, the convexity of the solution is not enforced as a constraint in (6). Instead, it arises implicitly from the structure of a Bellman operator.

(b) For monotone discretizations of Bellman equations there is a well-established framework of semismooth Newton methods, also known as Howard's algorithm [20], which guarantee global superlinear convergence when solving the finite-dimensional equation. These methods have a successful track record for large-scale computations. Howard's algorithm also ensures existence and uniqueness of numerical solutions.

(c) The treatment of the degenerate case $f(x) = 0$ is naturally incorporated in the converge proof and does not lead to complications in the analysis.

(d) The literature on numerical methods for Bellman-type equations is in various aspects richer than that for Monge–Ampère-type equations, for instance, because of the connection to stochastic control problems. As a result, one can use or adapt the numerical methods for Bellman-type equations to solve Monge–Ampère-type equations.

In order to permit unstructured meshes we employ continuous linear finite element spaces. Let $\mathcal{T}_h$ denote a shape-regular triangular or tetrahedral partition, where h is its mesh function. This means that

$$x \in T \text{ where } T \in \mathcal{T}_h \qquad \Longrightarrow \qquad h(x) = \operatorname{diam}(T). \tag{9}$$

On element boundaries $h(x)$ is equal to the diameter of the largest element neighboring it; so we could say that h is the upper semicontinuous function with domain $\overline{\Omega}$ satisfying (9). We abbreviate $\|h\|_{L^\infty(\Omega)}$ by $\mathbb{h}$. We denote by $\mathcal{N}_h^I$ and $\mathcal{N}_h^B$ respectively the interior and boundary grid points of $\mathcal{T}_h$ and set $\mathcal{N}_h := \mathcal{N}_h^I \cup \mathcal{N}_h^B$. The union of elements, denoted Ω_h, is called the computational domain. Because Ω is strictly convex, Ω_h cannot be equal to Ω. We require that Ω_h approximates Ω in the sense that $\mathcal{N}_h^B \subset \partial\Omega$ and $\Omega_h \subset \Omega$.

Let V_h denote the space of continuous piecewise linear polynomials over $\mathcal{T}_h$ and V_h^0 be the subspace of V_h consisting of those functions which vanish at every grid point in $\mathcal{N}_h^B$. Further, let $\{\psi_h^j\}_{j=1}^{J_0}$ denote the nodal basis for V_h^0 and $\{\psi_h^j\}_{j=1}^{J}$ denote the nodal basis for V_h, where $J_0 := \operatorname{card}(\mathcal{N}_h^I)$ and $J := \operatorname{card}(\mathcal{N}_h)$ are the cardinal numbers of $\mathcal{N}_h^I$ and $\mathcal{N}_h$, respectively. Often ψ_h^j is called a hat function. In order to study convergence of numerical solutions we need to embed V_h into $B(\Omega)$, i.e., extend the domain of $v \in V_h$ from Ω_h to Ω. We shall understand that $v \in V_h$ is extended as a constant along the outer normal vectors of $\partial\Omega_h$; see Figure 1. It is not intended that this extension is implemented in numerical codes.

We first state a basic finite difference formula, which serves as a building block for the numerical schemes in this paper. Let $\mathbf{b} \in \mathbf{R}^d$. For smooth $\phi : \mathbf{R}^d \to \mathbf{R}$ there holds for $k > 0$ and $x \in \mathbf{R}^d$

$$\begin{aligned} \operatorname{tr}\big[\mathbf{b}\mathbf{b}^T D^2\phi(x)\big] &= D^2\phi(x)\mathbf{b}\cdot\mathbf{b} = \partial_{\mathbf{b}\mathbf{b}}^2\phi(x) \\ &= \frac{\phi(x-k\mathbf{b}) - 2\phi(x) + \phi(x+k\mathbf{b})}{k^2} + O(k^2). \end{aligned} \tag{10}$$

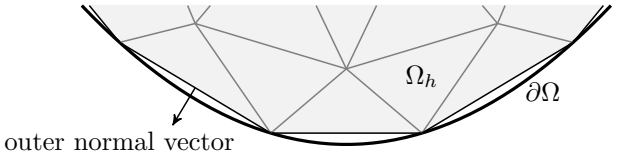


FIG. 1. *Ω is approximated by Ω_h so that the nodes on $\partial\Omega_h$ belong to $\partial\Omega$. To extend functions $v : \Omega_h \to \mathbf{R}$ to Ω, we assume that the extended function is constant along the normal coordinates of $\partial\Omega_h$ for $x \in \Omega \setminus \Omega_h$.*

The proof of (10) for $\phi \in C^4(\mathbf{R}^d)$ follows readily from an application of Taylor's formula. We omit the details.

For a $d \times d$ real valued matrix $\boldsymbol{\sigma}$, let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_d)$ with $\sigma_j \in \mathbf{R}^d$ denoting the jth column vector of $\boldsymbol{\sigma}$. Let $\boldsymbol{\sigma}^T$ be the transpose of $\boldsymbol{\sigma}$ and let $\boldsymbol{\lambda}$ be a diagonal matrix with λ_j in the jth position of the diagonal. Using (10) we immediately get for all $x \in \mathbf{R}^d$

$$\boldsymbol{\sigma}\boldsymbol{\lambda}\boldsymbol{\sigma}^T : D^2\phi(x) = \operatorname{tr}\left[\boldsymbol{\sigma}\boldsymbol{\lambda}\boldsymbol{\sigma}^T D^2\phi(x)\right] = \sum_{j=1}^{d} \operatorname{tr}\left[\lambda_j \sigma_j \sigma_j^T D^2\phi(x)\right]$$

$$= \sum_{j=1}^{d} \lambda_j \frac{\phi(x - k\sigma_j) - 2\phi(x) + \phi(x + k\sigma_j)}{k^2} + O(k^2), \tag{11}$$

where $A : B$ stands for the Frobenius inner product between two matrices A and B. It is an important feature of (11) that the explicit finite difference discretization of mixed derivatives is avoided in order to build monotonicity into the scheme.

The choice of k depends on h and x:

(a) It is known as the Wasow–Motzkin theorem [28, Theorem 1] that in order to achieve consistency with equations like (5a) and (6a) simultaneously with monotonicity, the mesh size h has to decrease locally strictly faster than the stencil size k; see also [22]. Therefore we expect k to decrease as the mesh size h shrinks, but within this "$h/k \to 0$" limitation. *In other words, the Wasow–Motzkin theorem implies that any monotone consistent method has to be a wide stencil scheme.*

(b) Observe that if $\phi \in C^4(\overline{\Omega})$, then (11) remains valid as long as the stencil size k is chosen small enough so that the stencil does not extend out of the domain. Hence near the boundary the size of k needs to be reduced to the size of h for $x \in \mathcal{N}_h^I$. This makes k dependent on x.

A specific choice for k is given in Remark 4.1 below. In general, condition (b) is reflected by the requirement that

$$k : \ L^\infty(\Omega) \times \Omega \to (0, \infty), \ (h, x) \mapsto k(h, x)$$

is a function such that $x - k(h,x)\,\sigma_j$ and $x + k(h,x)\,\sigma_j$ are in $\overline{\Omega}_h$ for all mesh functions h and $x \in \Omega_h$ and σ_j. Condition (a) is in conflict with this as (b) implies that h cannot decrease faster than k near $\partial\Omega$. Therefore we shall impose the Wasow–Motzkin limitation uniformly only on the subsets

$$\Omega_i = \left\{x \in \Omega : \operatorname{distance}(x, \partial\Omega) > \tfrac{1}{i}\right\},$$

illustrated in Figure 2; see also the related Figure 5. Thus on each Ω_i we require

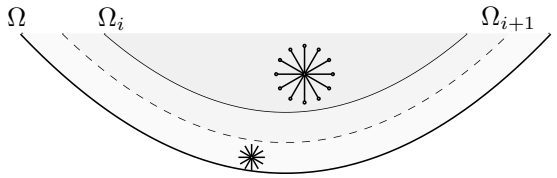


FIG. 2. *The Ω_i form a covering of Ω. On each of the Ω_i the Wasow–Motzkin consistency condition "$h/k \to 0$" is implemented uniformly; near the boundary this is not enforced as local stencils are rescaled so that they do not extend out of the computational domain—illustrated by two cartoon stencils in the figure.*

$$\sup_{x\in\Omega_i} \frac{h(x)}{k(h,x)} \to 0 \qquad \text{as} \qquad \mathbb{h} \to 0, \tag{12}$$

recalling that $\mathbb{h}$ is the largest diameter of an element of the mesh. Furthermore, we shall assume that on each Ω_i the stencil size k is eventually a constant function: for every $i \in \mathbf{N}$ there is an h' so that $x \mapsto k(h,x)$ is a constant function on Ω_i whenever $\|h\|_\infty < h'$. Moreover, we assume that the stencil size k shrinks uniformly, meaning that on the whole domain Ω

$$\sup_{x\in\Omega} k(h,x) \to 0 \qquad \text{as} \qquad \mathbb{h} \to 0. \tag{13}$$

Remark 4.1. As prototypical choice for k we have in mind that

$$k(h,x) = \min\{\kappa(\mathbb{h}), \text{distance}(x, \partial\Omega)\} \qquad \forall\, x \in \Omega \tag{14}$$

for some $\kappa : (0,\infty) \to (0,\infty)$ with $\xi/\kappa(\xi) \to 0$ and $\kappa(\xi) \to 0$ as $\xi \to 0$, e.g., $\kappa(\xi) = \sqrt{\xi}$. Observe that once $\kappa(\mathbb{h}) < \frac{1}{i}$ then $k = \kappa(\mathbb{h})$ is constant on the restriction to Ω_i because there $\kappa(\mathbb{h}) < \text{distance}(x, \partial\Omega)$. Since the calculation of distance$(x, \partial\Omega)$ can be computationally expensive, one should in practice implement an approximation of (14) satisfying (12) and (13).

To discretize the linear operators

$$\phi \mapsto -B : D^2\phi(x) + f\sqrt[d]{\det B}, \tag{15}$$

which are found under the supremum of (3), we choose factorizations $B = \boldsymbol{\sigma}\boldsymbol{\lambda}\boldsymbol{\sigma}^T$ for each $B \in \mathbf{S}_1$. More precisely, we consider some compact set

$$\mathbf{F} \subset \mathbf{R}^{d\times d} \times \{A \in \mathbf{R}^{d\times d} : A \text{ diagonal}\}$$

such that the mapping

$$\mathbf{F} \to \mathbf{S}_1,\ (\boldsymbol{\sigma}, \boldsymbol{\lambda}) \mapsto \boldsymbol{\sigma}\,\boldsymbol{\lambda}\,\boldsymbol{\sigma}^T \tag{16}$$

is bijective. Moreover we assume that all $\boldsymbol{\lambda}$ have the same trace:

$$\exists\, C > 0\, \forall\, (\boldsymbol{\sigma}, \boldsymbol{\lambda}) \in \mathbf{F} :\ \operatorname{tr}\boldsymbol{\lambda} = C. \tag{17}$$

The existence of such $\boldsymbol{\sigma}$ and $\boldsymbol{\lambda}$ follows from the symmetry of B. We remark that strictly speaking (16) only needs to be surjective for the subsequent analysis; however, without injectivity the notation becomes more cumbersome as more than one factorization represents a single B. We remark that our analysis also extends to direction dependent $k = k(h, x, \sigma_j)$, owing to the compactness of $\mathbf{F}$.

At this point there is considerable flexibility in the selection of $\mathbf{F}$. We discuss concrete choices in section 7, after examining the well-posedness of the discrete equations in section 5 and the convergence of numerical solutions in section 6.

The approximation of (15) is the mapping $L_h^B : \mathbf{R} \times B(\overline{\Omega}) \to B(\overline{\Omega})$, where for any $\phi \in B(\overline{\Omega})$ the value $L_h^B(s,\phi)(x_i)$ at internal node $x_i \in \mathcal{N}_h^I$ is set to be

$$-\sum_{j=1}^{d} \lambda_j \frac{\phi(x_i - k\sigma_j) - 2s + \phi(x_i + k\sigma_j)}{k^2} + f(x_i) \sqrt[d]{\det B}, \tag{18}$$

where $\boldsymbol{\sigma} = \boldsymbol{\sigma}(B)$ and $\boldsymbol{\lambda} = \boldsymbol{\lambda}(B)$ come from inversion of (16). Recall that $k = k(h,x)$ is chosen so that $x - k(h,x)\,\sigma_j$ and $x + k(h,x)\,\sigma_j$ are in $\overline{\Omega}_h$. Also observe how s takes, in comparison with (11), the place of $\phi(x_i)$. The auxiliary variable s allows us to express the monotonicity of the scheme efficiently in the language of the Barles–Souganidis framework [4], on which we model our proof of convergence. The value $L_h^B(s,\phi)(x_i)$ for boundary nodes $x_i \in \mathcal{N}_h^B$ is

$$s - g(x_i). \tag{19}$$

At nodes $x_i \in \mathcal{N}_h$, the Bellman operator H is represented approximately by

$$H_h(s,\phi)(x_i) = \sup_{B \in \mathbf{S}_1} L_h^B(s,\phi)(x_i).$$

For the remaining $x \in \Omega_h \setminus \mathcal{N}_h$ the value of $H_h(s,\phi)(x)$ is defined by piecewise linear interpolation of the nodal values, so that we have a mapping

$$H_h : \mathbf{R} \times B(\overline{\Omega}) \to B(\overline{\Omega}), \tag{20}$$

upon constant extension in the normal direction for $x \in \overline{\Omega} \setminus \Omega_h$; recall Figure 1.

Finally, our numerical scheme for (6) is defined as seeking $u_h \in V_h$ such that

$$H_h(u_h(x_i), u_h)(x_i) = 0 \qquad \forall\, x_i \in \mathcal{N}_h. \tag{21}$$

5. Well-posedness of the discrete equations. A common technique to show the well-posedness of a nonlinear system such as (21) is to formulate a fixed point argument akin to a pseudotime Euler scheme [13, 29]. However, to take advantage of the monotone discretization of the Bellman equation, we use instead Howard's algorithm [7, 20] to establish the existence and uniqueness of numerical solutions. This algorithm, being globally superlinearly converging, is also used to compute the numerical solutions of our numerical experiments in section 8.

Let

$$\mathbf{B} = (B_1, B_2, \ldots, B_{J_0}) = (\boldsymbol{\sigma}^{(1)}\boldsymbol{\lambda}^{(1)}(\boldsymbol{\sigma}^{(1)})^T, \boldsymbol{\sigma}^{(2)}\boldsymbol{\lambda}^{(2)}(\boldsymbol{\sigma}^{(2)})^T, \ldots, \boldsymbol{\sigma}^{(J_0)}\boldsymbol{\lambda}^{(J_0)}(\boldsymbol{\sigma}^{(J_0)})^T)$$

be an element of $\mathbf{S}_1^{J_0}$. Then $\mathsf{L}_h^{\mathbf{B}} : V_h \to V_h$ discretizes $\phi \mapsto -B_i : D^2\phi(x)$ at the internal nodes as

$$\mathsf{L}_h^{\mathbf{B}}(\phi)(x_i) = \begin{cases} -\sum_{j=1}^{d} \lambda_j^{(i)} \frac{\phi(x_i - k\sigma_j^{(i)}) - 2\phi(x_i) + \phi(x_i + k\sigma_j^{(i)})}{k^2} & : x_i \in \mathcal{N}_h^I, \\ \phi(x_i) & : x_i \in \mathcal{N}_h^B. \end{cases} \tag{22}$$

Similarly, we set

$$\mathsf{F}_h^{\mathbf{B}}(x_i) = \begin{cases} f(x_i)\sqrt[d]{\det B_i} & : \ x_i \in \mathcal{N}_h^I, \\ -g(x_i) & : \ x_i \in \mathcal{N}_h^B. \end{cases} \tag{23}$$

For the remaining $x \in \overline{\Omega_h} \setminus \mathcal{N}_h$ the values of $\mathsf{L}_h^{\mathbf{B}}(\phi)(x)$ and $\mathsf{F}_h^{\mathbf{B}}(x)$ are defined by piecewise linear interpolation of the nodal values. It is worthwhile to bring the differences between $\mathsf{L}_h^{\mathbf{B}}$ in (22) and L_h^B in (18) to mind. While the former has the right structure for the finite-dimensional analysis of Howard's algorithm, the latter mirrors the Barles–Souganidis formulation with the additional argument s to examine the monotonicity property efficiently.

LEMMA 5.1. *Let $\mathbf{B} \in \mathbf{S}_1^{J_0}$ and $h > 0$. If $\mathsf{L}_h^{\mathbf{B}} v \le 0$, then v attains its maximum at a boundary node. Moreover, the representation of the mapping $\mathsf{L}_h^{\mathbf{B}}$ as a matrix A, using the linear finite element hat functions as basis, is an invertible M-matrix.*

Proof. Let X be the set of nodes where v attains its maximum. Suppose that X consists only of internal nodes, i.e., $X \subset \mathcal{N}_h^I$, and that $\mathsf{L}_h^{\mathbf{B}} v \le 0$ holds. Let $C(X)$ be the convex hull of X. Let x_i be an extreme point of $C(X)$; it is clear that such x_i exists, not least by the Krein–Milman theorem, and that it is a node. For each $1 \le j \le d$, the value $v(x_i \pm k\sigma_j^{(i)})$ is a weighted average of the nodal values of v at the vertices x_ℓ of the finite element which contains $x_i \pm k\sigma_j^{(i)}$. It follows from $\mathsf{L}_h^{\mathbf{B}} v \le 0$ that v is equal to $v(x_i)$ at all those nodes x_ℓ whenever $\lambda_j \ne 0$, noting that there is at least one nonzero λ_j. Thus $x_i \pm k\sigma_j^{(i)} \in C(X)$, which contradicts that x_i is an extreme point. Hence X must contain a boundary node.

Suppose that $Av = 0$. Then v attains its maximum and, considering the argument of the previous paragraph for $-v$, its minimum on the boundary. As the restriction of A to boundary nodes is an identity map, it follows that $v = 0$. Hence A is invertible.

Owing to (22), $A \in \mathbf{R}^{n\times n}$ and $a_{ij} \le 0$ for all $i \ne j$ and with $n = J_0$. Moreover, $A + \epsilon\,\mathrm{Id}$ is strictly diagonally dominant for each $\epsilon > 0$. Therefore such $A + \epsilon\,\mathrm{Id}$ are M-matrices [5, Theorem (2.3) with (M_{35}), p. 137] and hence A is an M-matrix [5, Theorem (2.3) with (D_{15}), p. 135]. ∎

It follows directly from the construction of the discrete Hamiltonian that the numerical scheme, defined in (21), is equivalent to

$$\sup_{\mathbf{B}\in\mathbf{S}_1^{J_0}} \left[\mathsf{L}_h^{\mathbf{B}}(u_h)(x_i) + \mathsf{F}_h^{\mathbf{B}}(x_i)\right] = 0 \qquad \forall\, x_i \in \mathcal{N}_h.$$

For the solution of (21) we use Algorithm 1, known as Howard's method.

THEOREM 5.2. *Let $h > 0$ and assume that $\mathbf{F}$ is compact and (16) bijective. Then for every $g \in B(\partial\Omega)$ there exists a unique numerical solution $u_h \in V_h$ of (21). Moreover, the sequence $(v^\ell)_\ell$ generated by Howard's algorithm converges monotonically decreasing and superlinearly to u_h as $\ell \to \infty$.*

Algorithm 1. Howard's method.

1: Select an arbitrary $\mathbf{B}_0 \in \mathbf{S}_1^{J_0}$
2: **for** $\ell \in \mathbf{N}$ **do**
3: Let v^ℓ be the solution of the affine equation $\mathsf{L}_h^{\mathbf{B}}(v^\ell) + \mathsf{F}_h^{\mathbf{B}} = 0$.
4: Set $\mathbf{B}_{\ell+1} = \operatorname{argmax}_{\mathbf{B}\in\mathbf{S}_1^{J_0}} \left[\mathsf{L}_h^{\mathbf{B}}(v^\ell) + \mathsf{F}_h^{\mathbf{B}}\right]$
5: **end for**

Proof. Due to the bijectivity of (16) we may consider $\mathbf{F}$ instead of $\mathbf{S}_1$ as the set of controls. Clearly the mappings $(\boldsymbol{\sigma}, \boldsymbol{\lambda}) \mapsto \mathsf{L}_h^{\mathbf{B}}$ are $(\boldsymbol{\sigma}, \boldsymbol{\lambda}) \mapsto \mathsf{F}_h^{\mathbf{B}}$ are continuous. The monotonicity condition of [7] is verified above in Lemma 5.1. The result now follows from Theorem 2.1 of [7], noting that the substitution of the min in (1.1) of [7] by max reverses the direction of the monotone convergence. The superlinear rate follows from Theorem 3.4 of [7]. ∎

Observe that this well-posedness result for the discrete Bellman problem does not require convexity of the domain—the proof of Lemma 5.1 remains valid for nonconvex Ω, even though then possibly $C(X) \not\subset \Omega_h$, where $C(X)$ is defined in the proof of Lemma 5.1.

6. Convergence analysis. Comparison principles are a central component of the theory of viscosity solutions. With Perron's method they are commonly used to show existence of solutions. For the analysis of numerical methods, the Barles–Souganidis framework, which we loosely follow in this section, provides a convergence argument based on comparison of subsolutions and supersolutions.

Dirichlet boundary conditions pose here a particular difficulty. The strong comparison principle underlying the original Barles–Souganidis proof requires comparison of semicontinuous subsolutions and supersolutions, which obey boundary conditions in the viscosity sense. Yet, general degenerate elliptic equations usually only satisfy comparison of semicontinuous functions with pointwise Dirichlet conditions or comparison of continuous functions with viscosity Dirichlet conditions [12, section 7.C]. The combination as in the Barles–Souganidis framework without additional structure assumptions about the boundary value problem in general does not hold.

To resolve this mismatch we verify that the upper and lower semicontinuous envelopes of the numerical solutions satisfy the boundary conditions pointwise, at which point the Barles–Souganidis argument becomes in its essential steps available. In fact, this Lemma 6.4 is the only place in our analysis where the convexity of the domain is used, being aware that a Barles–Souganidis argument is a proof of existence and uniqueness of viscosity solutions.

We introduce

$$S:\ \mathbf{R}_+ \times \overline{\Omega} \times \mathbf{R} \times B(\overline{\Omega}) \to \mathbf{R},\ (\mathbb{h}, x, s, \phi) \mapsto \mathbb{h}\, H_h(s, \phi)(x), \tag{24}$$

to match precisely the structure of the solution operator in (2.1) of [4].

LEMMA 6.1. *The mapping S is monotone in the sense that*

$$S(\mathbb{h}, x, s, u) \le S(\mathbb{h}, x, s, v) \qquad \text{if} \qquad u \ge v$$

for all $\mathbb{h} > 0$, $x \in \overline{\Omega}$, $s \in \mathbf{R}$, and $u, v \in B(\overline{\Omega})$.

Proof. This follows directly from (18) and (19). ∎

For the proof of stability we construct a comparison function ζ. This ζ will subsequently also be used to show that the envelopes of the numerical solutions obey the boundary conditions in the pointwise sense; cf. Lemma 6.4.

LEMMA 6.2. *The mapping S is stable: there exists an h-independent constant $C > 0$ such that*

$$\|u_h\|_{L^\infty(\overline{\Omega})} \le C \tag{25}$$

for u_h given by (21)*. Furthermore, let $p \in \mathbf{R}^d$ and choose*

$$M \geq \|f\|_{L^\infty(\Omega)} \max_{B\in\mathbf{S}_1} \sqrt[d]{\det B} = \frac{\|f\|_{L^\infty(\Omega)}}{d}. \tag{26}$$

Let $\zeta(x;M,p) = \frac{M}{2}|x-p|^2$ and I_h be the nodal interpolant onto V_h. Then, for all $h>0$, the function $u_h - I_h\zeta$ (resp., $u_h + I_h\zeta$) attains its minimum (resp., maximum) over $\overline{\Omega}$ at a boundary node.

Proof. Let

$$\mathbf{B}' = \operatorname{argmax}_{\mathbf{B}\in\mathbf{S}_1^{J_0}} \left[\mathsf{L}_h^{\mathbf{B}}(u_h) + \mathsf{F}_h^{\mathbf{B}}\right].$$

Then $S(\mathbb{h}, x_i, u_h(x_i), u_h)/\mathbb{h} = \mathsf{L}_h^{\mathbf{B}'}(u_h)(x_i) + \mathsf{F}_h^{\mathbf{B}'}(x_i)$ at $x_i \in \mathcal{N}_h$.

To derive a bound on u_h from below, let $\zeta(x) = \zeta(x;M,p)$ be as in the statement of the theorem. Observe that for any internal node x_i, also near the boundary $\partial\Omega_h$,

$$\begin{aligned}\sum_{j=1}^d \lambda_j^{(i)} \frac{\zeta(x_i - k\sigma_j^{(i)}) - 2\zeta(x_i) + \zeta(x_i + k\sigma_j^{(i)})}{k^2} &= \sum_{j=1}^d \lambda_j \partial^2_{\sigma^{(i)},\sigma^{(i)}}\zeta(x_i) \\ &\overset{(11)}{=} B_i' : D^2\zeta(x_i) = M\,(B_i' : \mathrm{Id}) = M\,(\operatorname{tr} B_i') = M.\end{aligned}$$

Because of the convexity of ζ we know that

$$I_h\zeta(x_i - k\sigma_j^{(i)}) \geq \zeta(x_i - k\sigma_j^{(i)}), \qquad I_h\zeta(x_i + k\sigma_j^{(i)}) \geq \zeta(x_i + k\sigma_j^{(i)})$$

and consequently, since $I_h\rho(x_i) = \rho(x_i)$ as x_i is a node,

$$\sum_{j=1}^d \lambda_j^{(i)} \frac{I_h\zeta(x_i - k\sigma_j^{(i)}) - 2I_h\zeta(x_i) + I_h\zeta(x_i + k\sigma_j^{(i)})}{k^2} \geq M.$$

Hence, with $N \in \mathbf{R}$,

$$\mathsf{L}_h^{\mathbf{B}'}(I_h\zeta - u_h - N)(x_i) \leq \begin{cases} f(x_i)\sqrt[d]{\det B_i'} - M & : \text{ if } x_i \in \mathcal{N}_h^I, \\ \frac{M}{2}|x_i - p|^2 - g(x_i) - N & : \text{ if } x_i \in \mathcal{N}_h^B. \end{cases}$$

As, for N large, $\mathsf{L}_h^{\mathbf{B}'}(I_h\zeta - u_h - N) \leq 0$ on $\overline{\Omega}$ it follows from Lemma 5.1 that $I_h\zeta - u_h - N$ and equally $I_h\zeta - u_h$ attain their maximum at a boundary node x_i. Thus, for $x \in \overline{\Omega}$,

$$-u_h(x) \leq \|I_h\zeta\|_{L^\infty(\Omega)} + \|I_h\zeta - u_h\|_{L^\infty(\partial\Omega)} \leq 2\,\|\zeta\|_{L^\infty(\Omega)} + \|g\|_{L^\infty(\partial\Omega)}$$

gives an h-independent bound on u_h from below.

Now to the bound from above. As for large N we have $\mathsf{L}_h^{\mathbf{B}'}(u_h - N) \leq 0$ on $\overline{\Omega}$ it follows from Lemma 5.1 that u_h attains its maximum at a boundary node x_i, where $u_h(x_i) = g(x_i)$. Thus u_h is bounded from above by g. It is also clear that the maximizer of $u_h + I_h\zeta(x)$ is attained on $\partial\Omega_h$ and $\partial\Omega$, in fact for any $M \geq 0$. ∎

Our consistency condition (27) differs from [4] in that we require $x \in \Omega$ instead of $x \in \overline{\Omega}$; however, see also Lemma 6.4. Indeed we would not expect our scheme to be consistent as in the Barles–Souganidis framework, predicted by the results in [28, 22] due to the violation of (12) in the vicinity of the boundary. One would assume that any numerical method cropping a wide stencil near $\partial\Omega$ is incompatible with the original Barles–Souganidis framework, because the viscosity boundary conditions used there require the consistent monotone discretization of both the boundary operator and the differential operator at all $x \in \partial\Omega$ for all test functions ϕ; see (7.9)–(7.10) of [12] and (2.4) of [4].

LEMMA 6.3. *The mapping S of (24) is consistent in the sense that for all $x \in \Omega$ and $\phi \in C^4(\Omega)$ there hold*

$$\begin{aligned} \limsup_{\substack{\mathbb{h}\to 0\\ y\to x\\ \xi\to 0}} \frac{S(\mathbb{h}, y, \phi(y)+\xi, \phi+\xi)}{\mathbb{h}} &\le H(D^2\phi(x), f(x)),\\ \liminf_{\substack{\mathbb{h}\to 0\\ y\to x\\ \xi\to 0}} \frac{S(\mathbb{h}, y, \phi(y)+\xi, \phi+\xi)}{\mathbb{h}} &\ge H(D^2\phi(x), f(x)). \end{aligned} \tag{27}$$

Proof. There is an i such that $x \in \Omega_i$. Also $x \in \Omega_h$ for $\mathbb{h}$ sufficiently small. Recall that Ω_i and Ω_h are open. Hence we may restrict our attention to $y \in \Omega_i \cap \Omega_h$. Let $\mathbb{h}$ be small enough such that $k(h,x) + \mathbb{h} < \operatorname{dist}(y, \partial\Omega_h \cup \partial\Omega_i)$. The numerical operator fully expanded is

$$\begin{aligned} &\frac{1}{\mathbb{h}} S(\mathbb{h}, y, \phi(y)+\xi, \phi+\xi)\\ &= I_h\Big(x_i \mapsto \sup_{B\in\mathbf{S}_1} -\sum_{j=1}^{d} \lambda_j \frac{\phi(x_i + k\,\sigma_j) - 2\phi(y) + \phi(x_i - k\,\sigma_j)}{k^2} + f(x_i)\sqrt[d]{\det B}\Big)(y), \end{aligned}$$

where the interpolation operator I_h acts on a mapping which assigns to internal nodes x_i real values arising from the supremum over the finite-difference-like terms and the determinant term. The finite-difference-like terms depend on the y, however, and therefore are not central differences at this point. Finally, the interpolation operator returns an interpolating finite element function, which is evaluated at the very same y. The ξ, appearing in the Barles–Souganidis formulation (27) of consistency, cancels itself out immediately.

To prove consistency we first restore the central differences. We denote the B maximizing above at node x_i by $B_i = \boldsymbol{\sigma}^{(i)}\boldsymbol{\lambda}^{(i)}(\boldsymbol{\sigma}^{(i)})^T$. Then

$$\begin{aligned} &\frac{1}{\mathbb{h}} S(\mathbb{h}, y, \phi(y)+\xi, \phi+\xi)\\ &= I_h\Big(x_i \mapsto -\sum_{j=1}^{d} \lambda_j^{(i)} \frac{\phi(x_i + k\,\sigma_j^{(i)}) - 2\phi(y) + \phi(x_i - k\,\sigma_j^{(i)})}{k^2} + f(x_i)\sqrt[d]{\det B}\Big)(y)\\ &= I_h\Big(x_i \mapsto -\sum_{j=1}^{d} \lambda_j^{(i)} \frac{\phi(x_i + k\,\sigma_j^{(i)}) - 2\phi(x_i) + \phi(x_i - k\,\sigma_j^{(i)})}{k^2} + f(x_i)\sqrt[d]{\det B}\Big)(y)\\ &\quad + I_h\Big(x_i \mapsto \frac{2\,\operatorname{tr}(\boldsymbol{\lambda}^{(i)})\,\big(\phi(y) - \phi(x_i)\big)}{k^2}\Big)(y)\\ &\le I_h\Big(x_i \mapsto \sup_{B\in\mathbf{S}_1} -\sum_{j=1}^{d} \lambda_j \frac{\phi(x_i + k\,\sigma_j) - 2\phi(x_i) + \phi(x_i - k\,\sigma_j)}{k^2} + f(x_i)\sqrt[d]{\det B}\Big)(y)\\ &\quad + I_h\Big(x_i \mapsto \frac{2\,\operatorname{tr}(\boldsymbol{\lambda}^{(i)})\,\big(\phi(y) - \phi(x_i)\big)}{k^2}\Big)(y). \end{aligned}$$

Denoting the maximizing B in the last display at node x_i by $\bar{B}_i = \bar{\boldsymbol{\sigma}}^{(i)}\bar{\boldsymbol{\lambda}}^{(i)}(\bar{\boldsymbol{\sigma}}^{(i)})^T$, we obtain similarly

$$
\begin{aligned}
&\frac{1}{\mathbb{h}}\, S(\mathbb{h}, y, \phi(y)+\xi, \phi+\xi)\\
&\geq I_h\Big(x_i \mapsto -\sum_{j=1}^{d} \bar{\lambda}_j^{(i)} \frac{\phi(x_i + k\,\bar{\sigma}_j^{(i)}) - 2\phi(y) + \phi(x_i - k\,\bar{\sigma}_j^{(i)})}{k^2} + f(x_i)\sqrt[d]{\det B}\Big)(y)\\
&= I_h\Big(x_i \mapsto \sup_{B\in\mathbf{S}_1} -\sum_{j=1}^{d} \lambda_j \frac{\phi(x_i + k\,\sigma_j) - 2\phi(x_i) + \phi(x_i - k\,\sigma_j)}{k^2} + f(x_i)\sqrt[d]{\det B}\Big)(y)\\
&\quad + I_h\Big(x_i \mapsto \frac{2\,\operatorname{tr}(\bar{\boldsymbol{\lambda}}^{(i)})\,\big(\phi(y)-\phi(x_i)\big)}{k^2}\Big)(y).
\end{aligned}
$$

Because of (17) we conclude that the last two inequalities are in fact equalities and that the traces of $\boldsymbol{\lambda}$ and $\bar{\boldsymbol{\lambda}}$ may be taken out of I_h. For the test functions ϕ

$$
\sup_{z\in\Omega_i}\ \sup_{B\in\mathbf{S}_1}\ \sup_{1\leq j\leq d} \left| \frac{\phi(z + k\,\sigma_j) - 2\phi(z) + \phi(z - k\,\sigma_j)}{k^2} - \partial^2_{\sigma_j,\sigma_j}\phi(z) \right|
$$

is of the order $O(\sup_{z\in\Omega_i} k^2(h,z))$. Thence,

$$
\begin{aligned}
&I_h\Big(x_i \mapsto \sup_{B\in\mathbf{S}_1} -\sum_{j=1}^{d} \lambda_j \frac{\phi(x_i + k\,\sigma_j) - 2\phi(x_i) + \phi(x_i - k\,\sigma_j)}{k^2} + f(x_i)\sqrt[d]{\det B}\Big)(y)\\
&= I_h\Big(x_i \mapsto \sup_{B\in\mathbf{S}_1} -\sum_{j=1}^{d} \lambda_j \partial^2_{\sigma_j,\sigma_j}\phi(x_i) + f(x_i)\sqrt[d]{\det B}\Big)(y) + O(\sup_{z\in\Omega_i} k^2(h,z))\\
&\to H(D^2\phi(x), f(x))
\end{aligned}
$$

as $\mathbb{h}\to 0$ and $y\to x$, since $O(\sup_{z\in\Omega_i} k^2(h,z))\to 0$ as $\mathbb{h}\to 0$.

Finally we show that $I_h\big(x_i \mapsto (\phi(y)-\phi(x_i))/k^2\big)(y)\to 0$ as $\mathbb{h}\to 0$. Recall that

$$
\|\phi - I_h\phi\|_{L^\infty(\Omega_h)} \leq C\,\mathbb{h}^2 \|\phi\|_{W^{2,\infty}(\Omega_h)};
$$

see [14, Corollary 1.109]. There is a neighborhood N of y so that eventually all elements containing y belong to N and k is constant on N. Thus, for $\mathbb{h}$ small enough,

$$
I_h\Big(x_i \mapsto \frac{\phi(y)-\phi(x_i)}{k^2}\Big)(y) = \frac{\phi(y) - (I_h\phi)(y)}{k(h,y)^2} \to 0,
$$

as $\mathbb{h}\to 0$ due to (12). ∎

We define, with $x, y \in \overline{\Omega}$ and $h > 0$,

$$
\overline{u}(x) := \limsup_{\substack{y\to x\\ \mathbb{h}\to 0}} u_h(y) \qquad \text{and} \qquad \underline{u}(x) := \liminf_{\substack{y\to x\\ \mathbb{h}\to 0}} u_h(y). \tag{28}
$$

The following lemma confirms that $\overline{u}\in \mathrm{USC}(\overline{\Omega})$ and $\underline{u}\in \mathrm{LSC}(\overline{\Omega})$ are consistent with the pointwise Dirichlet conditions at the boundary.

LEMMA 6.4. *Let Ω be a strictly convex domain; then we have $\overline{u}(x) = \underline{u}(x) = g(x)$ for all $x\in\partial\Omega$.*

Proof. We show that $\underline{u}$ satisfies the pointwise boundary conditions on $\partial\Omega$. The proof for $\overline{u}$ is analogous. Fix $x\in\partial\Omega$. As Ω is convex there exists an affine mapping $P:\mathbf{R}^d\to\mathbf{R}$ such that

$$
(\overline{\Omega}\setminus\{x\}) \subset \{y\in\mathbf{R}^d : Py > 0\} \qquad \text{and} \qquad Px = 0.
$$

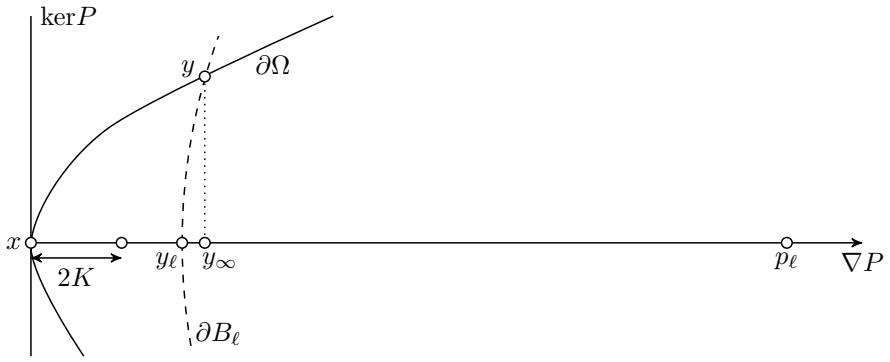


FIG. 3. *The constant K can be found for sufficiently large ℓ because y_ℓ converges to y_∞ as $\ell \to \infty$, where B_ℓ denotes the ball with center p_ℓ and radius $|y - p_\ell|$.*

For $\ell \in \mathbf{N}$ let $p_\ell = x + \ell \nabla P$, noting that ∇P is an inner normal vector of Ω. Again we use $\zeta_\ell(x) = \zeta(x; M, p_\ell)$ with M satisfying (26). We denote by q_ℓ the minimizer of $g - \zeta_\ell$ over $\partial\Omega$.

Due to compactness of $\partial\Omega$ the sequence $(q_\ell)_\ell$ has a converging subsequence with a limit $y \in \partial\Omega$. If $y \neq x$ it follows from the strict convexity that there is a constant $K > 0$ such that $|y - p_\ell| + 2K \leq |x - p_\ell|$ for all large ℓ; see Figure 3. But then, for $|q_\ell - y| < K$,

$$\zeta_\ell(x) = \frac{M}{2}|x - p_\ell|^2 > \frac{M}{2}\left(|q_\ell - p_\ell| + K\right)^2 = \zeta_\ell(q_\ell) + MK|q_\ell - p_\ell| + \frac{MK^2}{2}. \tag{29}$$

Because $C < MK|q_\ell - p_\ell|$ for large ℓ with C as in (25), (29) contradicts that q_ℓ is a minimizer. Hence

$$\lim_{\ell\to\infty} q_\ell = x. \tag{30}$$

Consider a sequence $(y_i, \mathbb{h}_i)_{i\in\mathbf{N}}$ with $\lim_{i\to\infty}(y_i, \mathbb{h}_i) = (x, 0)$. Then, for all $\ell \in \mathbf{N}$,

$$\begin{aligned}
\liminf_{i\to\infty} u_{h_i}(y_i) &= \lim_{i\to\infty} I_{h_i}\zeta_\ell(y_i) + \liminf_{i\to\infty}\left[u_{h_i}(y_i) - I_{h_i}\zeta_\ell(y_i)\right] \\
&\geq \zeta_\ell(x) + \liminf_{i\to\infty} \inf_{y\in\partial\Omega}\left[u_{h_i}(y) - I_{h_i}\zeta_\ell(y)\right] \\
&\geq \zeta_\ell(x) + \liminf_{i\to\infty} \inf_{y\in\partial\Omega}\left[g(y) - \zeta_\ell(y)\right] \\
&= \zeta_\ell(x) + g(q_\ell) - \zeta_\ell(q_\ell),
\end{aligned}$$

where we used that $u_{h_i} - I_{h_i}\zeta_\ell$ attains its minimum at a node on the boundary; cf. Lemma 6.2. Together with (30) we have

$$\liminf_{i\to\infty} u_{h_i}(y_i) \geq g(x).$$

As this inequality holds for all sequences $(y_i, \mathbb{h}_i)_{i\in\mathbf{N}}$ converging to $(x, 0)$, we have $\underline{u} \geq g$ on $\partial\Omega$. The opposite inequality follows by choosing sequences with $y_i = x$. ∎

We are ready to state the main result of this paper.

THEOREM 6.5. *Let Ω be a strictly convex domain. Assume that $f \in C(\Omega)$ with $f \geq 0$ and $g \in C(\partial\Omega)$. Then, as $\mathbb{h} \to 0$, the solutions u_h of (21) converge uniformly*

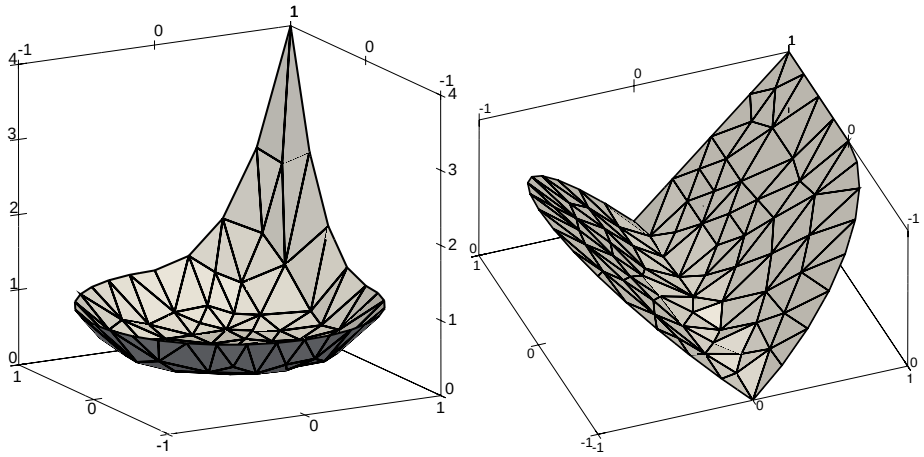
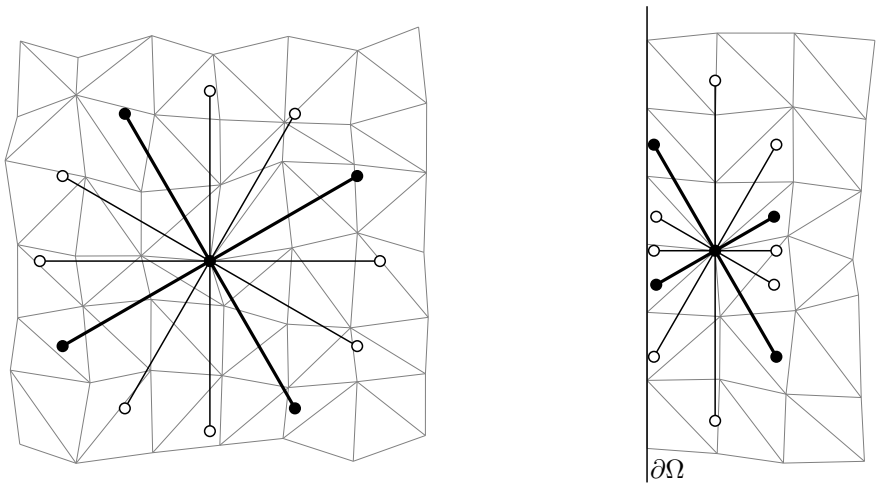


FIG. 4. *Solution of the quartic and nonsmooth problem on the coarsest mesh.*



(a) (b)

FIG. 5. *Plot* (a) *shows a stencil of the discrete Hamiltonian where the finite differences are spaced at angles of* $\pi/6$ *and* m *is about* 2.5. *The black dots mark a single stencil, the white dots stencil positions of other angles. Plot* (b) *illustrates how the finite differences are rescaled near the boundary to ensure that the stencil does not extend out of the boundary. We illustrate here how* $k = k(h, x, \sigma_j)$ *can also be rescaled depending on the direction* σ_j, *noting that our analysis easily extends to this case.*

to a function $u \in C(\overline{\Omega})$, *which is the unique viscosity solution on the set of convex functions of the Monge—Ampère problem* (5a). *This* u *is also the unique viscosity solution of the Bellman problem* (6a) *and it satisfies the Dirichlet boundary conditions pointwise.*

Proof. We have consistency for $\phi \in C^4(\Omega)$. It follows directly from the proof of Theorem 2.1 in [4] that for all $x \in \Omega$

$$H(D^2\phi(x), f(x)) \leq 0 \qquad (H(D^2\phi(x), f(x)) \geq 0)$$

quartic problem

DoFs	L^2-error	m	L^∞-error	m	H^1-error	m
91	$6.92 \cdot 10^{-2}$	2	$9.26 \cdot 10^{-2}$	2	$1.86 \cdot 10^{-1}$	2
329	$2.99 \cdot 10^{-2}$	2	$3.85 \cdot 10^{-2}$	2	$9.08 \cdot 10^{-2}$	2
1,249	$1.69 \cdot 10^{-2}$	4	$2.15 \cdot 10^{-2}$	2	$4.93 \cdot 10^{-2}$	4
4,865	$7.12 \cdot 10^{-3}$	4	$9.59 \cdot 10^{-3}$	4	$2.28 \cdot 10^{-2}$	4
19,201	$4.18 \cdot 10^{-3}$	8	$5.63 \cdot 10^{-3}$	4	$1.28 \cdot 10^{-2}$	8
76,289	$1.78 \cdot 10^{-3}$	8	$2.44 \cdot 10^{-3}$	8	$5.82 \cdot 10^{-3}$	8
304,129	$1.06 \cdot 10^{-3}$	16	$1.51 \cdot 10^{-3}$	8	$3.36 \cdot 10^{-3}$	8
1,214,465	$4.82 \cdot 10^{-4}$	16	$6.59 \cdot 10^{-4}$	16	$1.59 \cdot 10^{-3}$	16

nonsmooth problem

DoFs	L^2-error	m	L^∞-error	m	H^1-error	m
91	$4.50 \cdot 10^{-2}$	4	$1.03 \cdot 10^{-1}$	4	$2.02 \cdot 10^{-1}$	2
329	$1.62 \cdot 10^{-2}$	4	$5.69 \cdot 10^{-2}$	4	$1.51 \cdot 10^{-1}$	4
1,249	$7.11 \cdot 10^{-3}$	8	$3.08 \cdot 10^{-2}$	8	$1.21 \cdot 10^{-1}$	8
4,865	$3.35 \cdot 10^{-3}$	16	$2.03 \cdot 10^{-2}$	16	$9.80 \cdot 10^{-2}$	16
19,201	$1.70 \cdot 10^{-3}$	32	$1.38 \cdot 10^{-2}$	32	$7.91 \cdot 10^{-2}$	32
76,289	$9.63 \cdot 10^{-4}$	32	$9.12 \cdot 10^{-3}$	32	$6.33 \cdot 10^{-2}$	32
304,129	$5.10 \cdot 10^{-4}$	64	$6.04 \cdot 10^{-3}$	64	$5.12 \cdot 10^{-2}$	64
1,214,465	$3.12 \cdot 10^{-4}$	64	$4.51 \cdot 10^{-3}$	64	$4.25 \cdot 10^{-2}$	64

FIG. 6. *The second column shows the smallest relative L^2 error for a given grid across the factors $m \in \{2, 4, 8, 16, 32, 64\}$, with the minimizing m listed in the third column. The remaining columns are structured analogously.*

whenever $\underline{u} - \phi$ (resp., $\overline{u} - \phi$) attains a local maximum (minimum) at x. The result carries over [12, p. 57] to test functions $\phi \in C^2(\Omega)$ so that $\underline{u}$ and $\overline{u}$ are super- and subsolutions of (6a).

Now Lemmas 3.6 and 6.4 yield $\overline{u} \leq \underline{u}$ on $\overline{\Omega}$. The opposite inequality is clear from the definition of $\overline{u}$ and $\underline{u}$. This, together with (28), implies the uniform convergence to the unique viscosity solution of the Bellman problem on $\overline{\Omega}$. Now the result follows from Theorems 3.3 and 3.5. ∎

7. Parameter selection. It remains to show that a suitable compact set $\mathbf{F}$ can be found so that (16) is bijective. It turns out that there are several viable candidates.

A natural starting point is the eigen-decomposition $B = Q\Lambda Q^T$ of real symmetric matrices, where Q is an orthonormal matrix and Λ is a diagonal matrix whose entries are the eigenvalues of B. Similarly one can use the Cholesky decomposition or the closely related LDL decomposition $B = LDL^T$, where L is a lower unit triangular matrix and D a diagonal matrix. A widely used choice for the discretization of Bellman equations is $B = \boldsymbol{\sigma}\boldsymbol{\lambda}\boldsymbol{\sigma}^T = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$, that is, $\boldsymbol{\lambda} = \mathrm{Id}$; see [25] and section 5 of [13].

From the implementational point of view it is desirable to keep the set of $\boldsymbol{\sigma}$ small: While evaluation of $\phi(x_i - k\sigma_j^{(i)})$ and $\phi(x_i + k\sigma_j^{(i)})$ in (22) can be implemented efficiently [1, Remark 4], unnecessary evaluations should be avoided, especially if $k \gg h$. In contrast no significant savings arise in (22) from a small set of $\boldsymbol{\lambda}$. In the numerical experiments in the next section we use therefore the eigen-decomposition of B as in this case the σ_j can be normalized and multiple B share the same $\boldsymbol{\sigma} = Q$.

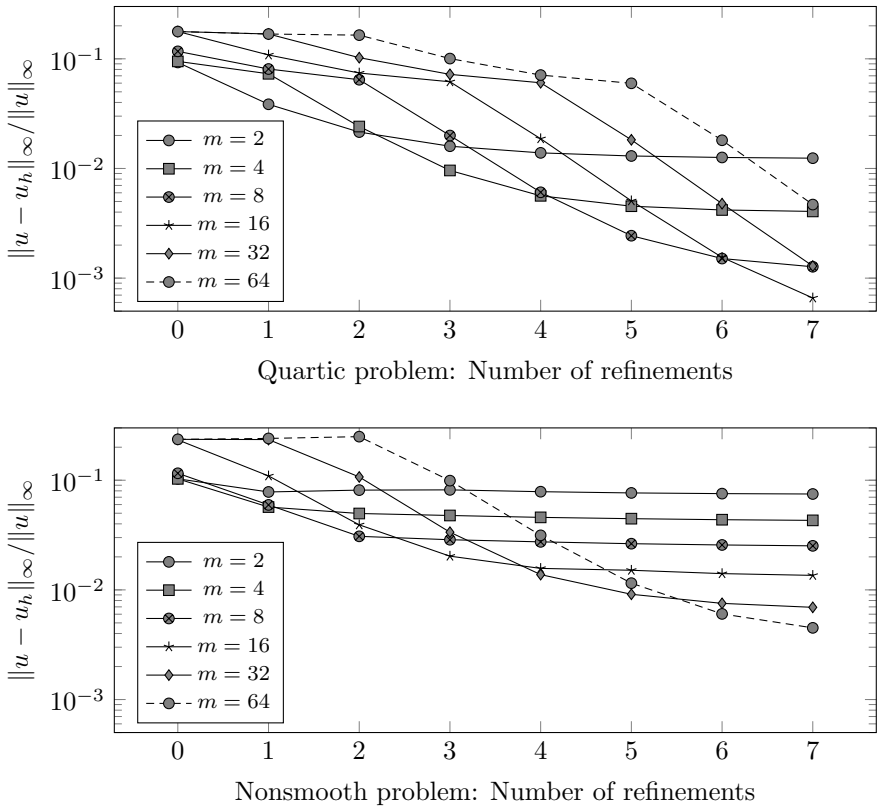


FIG. 7. *Relative L^∞-error for the test problem with quartic (above) and nonsmooth (below) exact solution.*

	m for quartic problem						m for nonsmooth problem					
refinement	2	4	8	16	32	64	2	4	8	16	32	64
0	**5**	5	5	4	5	5	4	**5**	6	5	5	5
1	**5**	5	6	10	5	5	4	**5**	6	7	9	6
2	**5**	5	7	9	12	5	5	5	**6**	6	7	11
3	5	**6**	7	9	12	13	5	5	7	**7**	8	9
4	5	**6**	7	11	12	16	7	5	6	7	**7**	7
5	6	6	**6**	10	12	15	8	6	6	7	**8**	8
6	5	6	**6**	9	12	15	7	6	6	7	8	**8**
7	5	5	7	**8**	10	14	8	5	7	7	8	**9**

FIG. 8. *Number of Newton iterations to achieve a Newton step size of less than 10^{-6}. The boxes highlight the factor m which minimizes the L^∞-error for a given level of refinement.*

8. Numerical experiments. In this section we present two two-dimensional numerical experiments to test the proposed wide-stencil method and Howard's Newton solver. The first experiment has the exact smooth solution $u(x) = |x|^4 = (x_1^2 + x_2^2)^2$ and the second experiment computes the nonsmooth viscosity solution $u(x) = |x_1|$. In both experiments the computational domain is the union of the unit circle and the

unit square so that the strict convexity condition is violated in part of the domain:

$$\Omega = \{(x,y) \in \mathbf{R}^2 : x^2 + y^2 < 1\} \cup \{(x,y) \in \mathbf{R}^2 : 0 < x, y < 1\}.$$

The quasi-uniform grid has at the coarsest level 91 nodes and at the finest level after 7 uniform refinements 1,214,465 nodes. The computations were carried out in Python with FEniCS on an Apple iMac computer. The numerical solutions on the coarsest grid are shown in Figure 4.

The compact control set is

$$\mathbf{F} = \left(\mathrm{SO}(2) \times \left\{ \begin{pmatrix} a & 0 \\ 0 & 1-a \end{pmatrix} : a \in \left[0, \tfrac{1}{2}\right) \right\}\right) \cup \left\{ \left(\mathrm{Id}, \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}\right) \right\}.$$

In order to compute the numerical solutions we discretize the special orthogonal group $\mathrm{SO}(2)$ by considering only the rotation angles $i\pi/64$, $i \in \{0, 1, \ldots, 63\}$; see Figure 5(a) for an illustration of angles $i\pi/6$. The stencil diameter k is, away from the boundary, represented through $k = m \cdot h$ by a fixed positive factor m and the (average) mesh size h. Near the boundary, so where $m \cdot h$ is larger than the distance to $\partial\Omega_h$, the stencil is reduced in size to remain within Ω_h; see Figure 5(b).

The relative errors in the L^2, L^∞, and H^1 norms when approximating the quartic and nonsmooth exact solution are summarized in Figure 6. The L^∞-error graphs for different m are plotted in Figure 7. Across the seven levels of refinement the orders of convergence in h and k are, with C representing generic constants,

	quartic problem		nonsmooth problem	
$\|u - u_h\|_{L^2}$	$\approx Ch^{1.02}$	$\approx Ck^{1.79}$	$\approx Ch^{1.02}$	$\approx Ck^{2.39}$
$\|u - u_h\|_{L^\infty}$	$\approx Ch^{1.02}$	$\approx Ck^{1.78}$	$\approx Ch^{0.64}$	$\approx Ck^{1.50}$
$\|u - u_h\|_{H^1}$	$\approx Ch^{0.98}$	$\approx Ck^{1.72}$	$\approx Ch^{0.33}$	$\approx Ck^{1.22}$

The number of Newton iterations in Figure 8 increases only moderately with the level of refinement and stencil size, so that fine meshes remain feasible on desktop computers. Importantly, Howard's algorithm displays a robust performance when approximating the nonsmooth solution $|x_1|$ with $f = 0$; noting that the line $\{x_1 = 0\}$ where $|x_1|$ is nondifferentiable is not aligned with the computational mesh. The iterations are started with the control $\mathbf{B}_0 = \frac{1}{d}\mathrm{Id}$. Due to global convergence, the starting iterate does not need to be guessed in close vicinity of the numerical solution. The stopping criterion is an iteration step size less than 10^{-6} in the L^∞-norm.

Acknowledgment. The authors would like to thank the referees for their careful reading and for their critical questions, which helped to improve the paper.

REFERENCES

[1] Y. Achdou and M. Falcone, *A semi-Lagrangian scheme for mean curvature motion with nonlinear Neumann conditions*, Interfaces Free Bound., 14 (2012), pp. 455–485.

[2] N. E. Aguilera and P. Morin, *On convex functions and the finite element method*, SIAM J. Numer. Anal., 47 (2009), pp. 3139–3157.

[3] O. Alvarez, J.-M. Lasry, and P.-L. Lions, *Convex viscosity solutions and state constraints*, J. Math. Pures Appl., 76 (1997), pp. 265–288.

[4] G. Barles and P. E. Souganidis, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptot. Anal., 4 (1991), pp. 271–283.

[5] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Classics in Appl. Math., SIAM, Philadelphia, 1994.

[6] J. F. Bonnans and H. Zidani, *Consistency of generalized finite difference schemes for the stochastic HJB equation*, SIAM J. Numer. Anal., 41 (2003), pp. 1008–1021.
[7] O. Bokanowski, S. Maroso, and H. Zidani, *Some convergence results for Howard's algorithm*, SIAM J. Numer. Anal., 47 (2009), pp. 3001–3026.
[8] S. C. Brenner, T. Gudi, M. Neilan, and L.-Y. Sung, *C^0 penalty methods for the fully nonlinear Monge-Ampère equation*, Math. Comp., 80 (2011), pp. 1979–1995.
[9] L. A. Caffarelli and X. Cabré, *Fully Nonlinear Elliptic Equations*, AMS, Providence, RI, 1995.
[10] F. Camilli and M. Falcone, *An approximation scheme for the optimal control of diffusion processes*, RAIRO Anal. Numer., 29 (1995), pp. 97–122.
[11] F. Camilli and E. R. Jakobsen, *A finite element like scheme for integro-partial differential Hamilton-Jacobi-Bellman equations*, SIAM J. Numer. Anal., 47 (2009), pp. 2407–2431.
[12] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc., 27 (1992), pp. 1–67.
[13] K. Debrabant and E. R. Jakobsen, *Semi-Lagrangian schemes for linear and fully nonlinear diffusion equations*, Math. Comp., 82 (2012), pp. 1433–1462.
[14] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Springer, New York, 2004.
[15] X. Feng, R. Glowinski, and M. Neilan, *Recent developments in numerical methods for second order fully nonlinear partial differential equations*, SIAM Rev., 55 (2013), pp. 205–267.
[16] X. Feng, C. Kao, and T. Lewis, *Convergent finite difference methods for one-dimensional fully nonlinear second order partial differential equations*, J. Comput. Appl. Math., 254 (2014), pp. 81–98.
[17] X. Feng and M. Neilan, *Mixed finite element methods for the fully nonlinear Monge–Ampére equation based on the vanishing moment method*, SIAM J. Numer. Anal., 47 (2009), pp. 1226–1250.
[18] D. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*, Springer, Berlin, 2001.
[19] C. E. Gutiérrez, *The Monge-Ampère Equation*, Birkhäuser, Basel, 2001.
[20] R. A. Howard, *Dynamic Programming and Markov Processes*, MIT Press, Cambridge, MA, 1960.
[21] M. Jensen and I. Smears, *On the convergence of finite element methods for Hamilton-Jacobi-Bellman equations*, SIAM J. Numer. Anal., 51 (2013), pp. 137–162.
[22] M. Kocan, *Approximation of viscosity solutions of elliptic partial differential equations on minimal grids*, Numer. Math., 72 (1995), pp. 73–92.
[23] N. V. Krylov, *Nonlinear Elliptic and Parabolic Equations of the Second Order*, Springer, New York, 1987.
[24] N. V. Krylov, *Fully nonlinear second order elliptic equations: recent development*, Ann. Sc. Norm. Super. Pisa Cl. Sci., 25 (1997), pp. 569–595.
[25] H. J. Kushner, *Numerical methods for stochastic control problems in continuous time*, SIAM J. Optim., 28 (1990), pp. 999–1048.
[26] O. Lakkis and T. Pryer, *A finite element method for nonlinear elliptic problems*, SIAM J. Sci. Comput., 35 (2013), pp. A2025–A2045.
[27] P. L. Lions, *Two remarks on Monge–Ampère equations*, Ann. Mat. Pure Appl., 142 (1985), pp. 262–275.
[28] T. S. Motzkin and W. Wasow, *On the approximation of linear elliptic differential equations by difference equations with positive coefficients*, J. Math. Phys., 31 (1953), pp. 253–259.
[29] A. M. Oberman, *Wide stencil finite difference schemes for the elliptic Monge–Ampère equation and functions of the eigenvalues of the Hessian*, Discrete Contin. Dyn. Syst. Ser. B, 10 (2008), pp. 221–238.
[30] I. Smears and E. Süli, *Discontinuous Galerkin finite element approximation of Hamilton–Jacobi–Bellman equations with Cordes coefficients*, SIAM J. Numer. Anal., 52 (2014), pp. 993–1016.